Medidas de Dispersión

Esta guía contiene actividades y preguntas que deben realizarse en sus cuadernos.
Una vez que las actividades presenciales se reanuden, serán revisadas todas las actividades que hayan sido solicitadas.

OA2: Tomar decisiones en situaciones de incerteza que involucren el análisis de datos estadísticos con medidas de dispersión y probabilidades condicionales.
OAc: Tomar decisiones fundamentadas en evidencia estadística y/o en la evaluación de resultados obtenidos a partir de un modelo probabilístico.
OAd: Argumentar, utilizando lenguaje simbólico y diferentes representaciones, para justificar la veracidad o falsedad de una conjetura, y evaluar el alcance y los límites de los argumentos utilizados.

Al igual que las medidas de tendencia central (media, moda, mediana), las medidas de dispersión son valores que se pueden calcular a partir de un conjunto de datos y que permiten resumir algunas de sus características a través de un valor numérico.

A diferencia de las medidas de tendencia central, las medidas de dispersión nos aportan información sobre qué tan alejados se encuentran nuestros datos los unos de los otros, información que puede ser utilizada para tomar decisiones en múltiples situaciones en que el nivel de dispersión de la información sea fundamental.

La medida de dispersión por excelencia es la llamada **Desviación Estándar**, simbolizada generalmente con la letra griega *sigma* minúscula σ. Pero antes de llegar a definirla y estudiar su utilidad, debemos tomar un recorrido por las distintas medidas de dispersión que existen.

Medidas de Dispersión

1. **Rango**: El rango es la diferencia (resta) que existe entre el dato de mayor valor y el de menor valor en un conjunto de datos.

Ejemplo
Determinar el rango del siguiente conjunto de datos

1	2	3	2	3	1	0	4	5
6	2	3	4	5	7	1	4	5
6	7	3	1	2	4	5	-1	-3
3	3	2	-4	4	-4	2	-5	2

Primero identificamos el mayor y el menor de los valores
El mayor es 7
El menor es -5

Luego, el rango se calcula restando el mayor menos el menor, en este caso

$$\text{Rango} = 7 - -5 = 12$$

2. **Desviación**: Es la diferencia (resta) que existe entre un dato del conjunto y la media de todos los datos del conjunto.

Ejemplo
Determinar la desviación de cada dato del siguiente conjunto

10	11	12	9	8
8	13	7	11	11

Dado que calcular la desviación requiere conocer la media, entonces es lo primero que debemos calcular. En este caso, la media es 10. Luego, la desviación de cada dato será

Dato	Operación	Desviación
7	7-10	-3
8	8-10	-2
9	9-10	-1
10	10-10	0
11	11-10	1
12	12-10	2
13	13-10	3

Se dice entonces, por ejemplo, que la desviación del 9 es -1.

3. **Desviación Media**: Es, como lo indica su nombre, el promedio de las desviaciones individuales de todos los datos de una muestra. Sin embargo, para que su valor no sea cero en todos los conjuntos de datos, es necesario que para realizar el cálculo se consideren los **valores absolutos** de las desviaciones de cada dato.

Ejemplo
Determinar la desviación media del conjunto de datos del ejemplo anterior.

Como ya hemos calculado las desviaciones de los datos, ahora solo necesitamos tener dos consideraciones. Primero, que hay datos que aparecen más de una vez, y segundo, que debemos tomar los valores absolutos de sus desviaciones.

Observa la siguiente tabla, en ella se han añadido columnas adicionales para atender estas consideraciones

Dato	Operación	Desviación	Frecuencia	Valor absoluto del aporte por dato
7	7-10	-3	1	3
8	8-10	-2	2	4
9	9-10	-1	1	1
10	10-10	0	1	0
11	11-10	1	3	3
12	12-10	2	1	2
13	13-10	3	1	3
Total			10	16

En la última columna se ubican los valores absolutos de los resultados de multiplicar las desviaciones por las frecuencias de cada dato, de modo que al sumar los valores de esta columna se obtenga *la suma total de los valores absolutos de las desviaciones de todos los datos del conjunto*.
Finalmente tenemos que calcular el promedio, por lo tanto, debemos dividir *la suma total de los valores absolutos de las desviaciones de todos los datos del conjunto* entre la cantidad total de datos.

$$\text{Desviación Media} = \frac{16}{10} = 1{,}6$$

Usted puede verificar que, de no haber considerado los valores absolutos de las desviaciones, el 16 hubiese sido 0.

4. **Varianza**: Semejante a la desviación media, la varianza es un promedio de las distancias que existen entre los datos del conjunto y su media. Sin embargo, en este caso la distancia ya no se medirá como la diferencia o resta entre ellos, sino a partir del cuadrado de dicha diferencia. Luego, la varianza es el promedio de los cuadrados de las desviaciones.

Ejemplo
Determinar la varianza del mismo conjunto de datos del ejemplo anterior.

En nuestra tabla, debemos agregar ahora una columna para calcular el cuadrado de las desviaciones.

Observa la siguiente tabla, en ella se han añadido columnas adicionales para atender estas consideraciones

Dato	Operación	Desviación	Cuadrado de la Desviación	Frecuencia	Aporte por dato
7	7-10	-3	9	1	9
8	8-10	-2	4	2	8
9	9-10	-1	1	1	1
10	10-10	0	0	1	0
11	11-10	1	1	3	3
12	12-10	2	4	1	4
13	13-10	3	9	1	9
Total				10	34

Nota que ya no hace falta especificar que se toman valores absolutos, dado que los cuadrados siempre serán valores no negativos.

Al igual que en el ejemplo anterior, debemos terminar haciendo la división entre la suma de los valores de la última columna con la cantidad de datos, así

$$\text{Varianza} = \frac{34}{10} = 3{,}4$$

5. **Desviación Estándar**: Es la raíz cuadrada de la varianza.

Resumen

Rango	es la diferencia entre los valores extremos
Desviación	es la diferencia de cada dato con la media
Desviación Media	es el promedio de los valores absolutos de todas las desviaciones
Varianza	es el promedio de los cuadrados de todas las desviaciones
Desviación Estándar	es la raíz cuadrada de la varianza

Ejemplo General
Determinar las medidas de dispersión del siguiente conjunto de datos

-1	2	0	2	3
-2	4	2	5	3
-2	4	0	-3	2
2	1	1	-2	-1

El Rango del conjunto de datos es

$$\text{Rango} = 5 - -3 = 8$$

Realizaremos una tabla para calcular las demás medidas de dispersión

Para abreviar la tabla se utilizarán símbolos en lugar de palabras, el más importante de ellos, el que representa la media del conjunto $\bar{x}$, cuyo valor es 1 en este caso.

Datos	Valor Absoluto de la Desviación	Cuadrado de la Desviación	Frecuencia	Aporte para la Desviación Media	Aporte para la Varianza

x	$\lvert x-\bar{x}\rvert$	$(x-\bar{x})^2$	f	$\mathrm{f}\cdot\lvert x-\bar{x}\rvert$	$\mathrm{f}\cdot(x-\bar{x})^2$
-3	4	16	1	4	16
-2	3	9	3	9	27
-1	2	4	2	4	8
0	1	1	2	2	2
1	0	0	2	0	0
2	1	1	5	5	5
3	2	4	2	4	8
4	3	9	2	6	18
5	4	16	1	4	16
Total			20	38	100

Así, gracias a la tabla podemos concluir rápidamente

La desviación media del conjunto es

$$\frac{38}{20} = 1{,}7$$

La varianza del conjunto es

$$\frac{100}{20} = 5$$

La desviación estándar del conjunto es

$$\sqrt{5} \approx 2.236$$

Ejercicios: Determinar las Medidas de Dispersión de los siguientes conjuntos de datos construyendo una tabla semejante a la vista en los ejemplos.

Conjunto 1

10	11	11	8	7
13	8	8	14	10

Conjunto 2

-1	-1	3	2	0
-2	2	1	1	-3
-3	-2	1	-1	0
-2	0	1	2	3

Conjunto 3

2	1	2	1	3	3	2	1	4	3
1	2	2	1	2	1	1	1	2	2
2	2	0	2	3	1	1	1	0	4
3	1	0	-3	3	-2	0	-2	-1	2
0	-1	1	-2	0	-2	1	-4	0	3